

SPECIFICATION

Title of the Invention

Process for Producing Microbial Transglutaminase

5

Background of the Invention

The present invention relates to a protein having a transglutaminase activity, DNA which encodes for the protein, and a process for producing the protein. In particular, the present invention relates to a process for producing a protein having a transglutaminase activity by a genetic engineering technique.

Transglutaminase is an enzyme which catalyzes the acyl transfer reaction of a γ -carboxyamido group in a peptide chain of a protein. When such an enzyme reacts with the protein, a reaction of an ϵ -(γ -Glu)-Lys forming reaction or substitution reaction of Gln with Glu by the deamidation of Glu can occur.

The transglutaminase is used for the production of gelled foods such as jellies, yogurts, cheeses, gelled cosmetics, etc. and also for improving the quality of meats [see Japanese Patent Publication for Opposition Purpose (hereinafter referred to as "J. P. KOKOKU") No. Hei 1-50382]. The transglutaminase is also used for the production of a material for microcapsules having a high thermal stability and a carrier for an immobilized enzyme. The transglutaminase is thus industrially very useful.

25

As for transglutaminases, those derived from animals and those derived from microorganisms (microbial transglutaminase; hereinafter

referred to as "MTG") have been known hitherto.

The transglutaminases derived from animals are calcium ion-dependent enzymes which are distributed in organs, skins and bloods of animals. They are, for example, guinea pig liver transglutaminase [K. Ikura et al., Biochemistry 27, 2898 (1988)], human epidermis keratin cell transglutaminase [M. A. Philips et al., Proc. Natl. Acad. Sci. USA 87, 9333 (1990)] and human blood coagulation factor XIII (A. Ichinose et al., Biochemistry 25, 6900 (1990)).

As for the transglutaminases derived from microorganisms, those independent on calcium were obtained from microorganisms of the genus Streptovorticillium. They are, for example, Streptovorticillium griseocarneum IFO 12776, Streptovorticillium cinnamoneum sub sp. cinnamoneum IFO 12852 and Streptovorticillium mobaraense IFO 13819 [see Japanese Patent Unexamined Published Application (hereinafter referred to as "J. P. KOKAI") No. Sho 64-27471].

According to the peptide mapping and the results of the analysis of the gene structure, it was found that the primary structure of the transglutaminase produced by the microorganism is not homology with that derived from the animals at all (European Patent publication No. 0 481 504 A1).

Since the transglutaminases (MTG) derived from microorganisms are produced by the culture of the above-described microorganisms followed by the purification, they had problems in the supply amount, efficiency, and the like. It is also tried to produce them by a genetic engineering technique. This technique includes a process which is conducted by the secretion expression of a microorganism such as E.

coli, yeast or the like (J. P. KOKAI NO. Hei 5-199883), and a process wherein MTG is expressed as an inactive fusion protein inclusion body in E. coli, this inclusion body is solubilized with a protein denaturant, the denaturant is removed and then MTG is reactivated to obtain the active MTG (J. P. KOKAI No. Hei 6-30771).

However, these processes have problems when they are practiced on an industrial scale. Namely, when the secretion by the microorganisms such as E. coli and yeast is employed, the amount of the product is very small; and when MTG is obtained in the form of the inactive fusion protein inclusion body in E. coli, an expensive enzyme is necessitated for the cleavage.

It is known that when a foreign protein is secreted by the genetic engineering method, the amount thereof thus obtained is usually small. On the contrary, it is also known that when the foreign protein is produced in the cell of E. coli, the product is in the form of inert protein inclusion body in many cases although the expressed amount is high. The protein inclusion body must be solubilized with a denaturant, the denaturing agent must be removed and then MTG must be reactivated.

It is already known that in the expression in E. coli, an N-terminal methionine residue in natural protein obtained after the translation of gene is efficiently cleaved with methionine aminopeptidase. However, the N-terminal methionine residue is not always cleaved in an exogenous protein.

Processes proposed hitherto for obtaining a protein free from N-terminal methionine residue include a chemical process wherein a protein

having methionine residue at the N-terminal or a fusion protein having a peptide added thereto through methionine residue is produced and then the product is specifically decomposed at the position of methionine residue with cyanogen bromide; and an enzymatic process wherein a recognition sequence of a certain site-specific proteolytic enzyme is inserted between a suitable peptide and an intended peptide to obtain a fusion peptide, and the site-specific hydrolysis is conducted with the enzyme.

However, the former process cannot be employed when the protein sequence contains a methionine residue, and the intended protein might be denatured in the course of the reaction. The latter process cannot be employed when a sequence which is easily broken down is contained in the protein sequence because the yield of the intended protein is reduced. In addition, the use of such a proteolytic enzyme is unsuitable for the production of protein on an industrial scale from the viewpoint of the cost.

Conventional processes for producing MTG have many problems such as supply amount and cost. Namely, in the secretion expression by *E. coli*, yeast or the like, the expressed amount is disadvantageously very small. In the production of the fusion protein inclusion body in *E. coli*, it is necessary, for obtaining mature MTG, to cleave the fusion part with restriction protease after the expression. Further, it has been found that since MTG is independent on calcium, the expression of active MTG in the cell of a microorganism is fatal because this enzyme acts on the endoprotein.

Thus, for the utilization of MTG, produced by the gene

recombination, on an industrial scale, it is demanded to increase the production of mature MTG free of the fusion part. The present invention has been completed for this purpose. The object of the present invention is to product MTG in a large amount in microorganisms such as E. coli.

When MTG is expressed with recombinant DNA of the present invention, methionine residue is added to the N-terminal of MTG. However, by the addition of the methionine residue to the N-terminal of MTG, there is some possibility wherein problems of the safety such as impartation of antigenicity to MTG occur. It is another problem to be solved by the present invention to produce MTG free of methionine residue corresponding to the initiation codon.

Summary of the Invention

An object of the present invention is to provide a novel protein having a transglutaminase activity.

Another object of the present invention is to provide a DNA encoding for the novel protein having a transglutaminase activity.

Another object of the present invention is to provide a recombinant DNA encoding for the novel protein having a transglutaminase activity.

Another object of the present invention is to provide a transformant obtained by the transformation with the recombinant DNA.

Another object of the present invention is to provide a process for producing a protein having a transglutaminase activity.

These and other objects of the present invention will be

apparent from the following description and examples.

For solving the above-described problems, the inventors have constructed a massive expression system of protein having transglutaminase activity by changing the codon to that for *E. coli*, or preferably by using a multi-copy vector (pUC19) and a strong promoter (*trp* promoter).

Since MTG is expressed and secreted in the prepro-form from microorganisms of actinomycetes, the MTG does not have methionine residue corresponding to the initiation codon at the N-terminal, but the protein expressed by the above-described expression method has the methionine residue at the N-terminal thereof. To solve this problem, the inventors have paid attention to the substrate specificity of methionine aminopeptidase of *E. coli*, and succeeded in obtaining a protein having transglutaminase activity and free from methionine at the N-terminal by expressing the protein in the form free from the aspartic acid residue which is the N-terminal amino acid of MTG. The present invention has been thus completed.

Namely, the present invention provides a protein having a transglutaminase activity, which comprises a sequence ranging from serine residue at the second position to proline residue at the 331st position in an amino acid sequence represented by SEQ ID No. 1 wherein N-terminal amino acid of the protein corresponds to serine residue at the second position of SEQ ID No. 1.

There is provided a protein which consists of an amino acid sequence of from serine residue at the second position to proline residue at the 331st position in an amino acid sequence of SEQ ID No. 1.

There is provided a DNA which codes for said proteins.

There is provided a recombinant DNA having said DNA, in particular, a recombinant DNA expressing said DNA.

There is provided a transformant obtained by the transformation with the recombinant DNA.

There is provided a process for producing a protein having a transglutaminase activity, which comprises the steps of culturing the transformant in a medium to produce the protein having a transglutaminase activity and recovering the protein.

Taking the substrate specificity of methionine aminopeptidase into consideration, the process for producing the protein having transglutaminase activity and free of initial methionine is not limited to the removal of the N-terminal aspartic acid.

Brief Explanation of the Drawings

Fig. 1 shows a construction scheme of MTG expression plasmid pTRPMTG-01.

Fig. 2 shows a construction scheme of MTG expression plasmid pTRPMTG-02.

Fig. 3 is an expansion of SDS-polyacrylamide electrophoresis showing that MTG was expressed.

Fig. 4 shows a construction scheme of MTG expression plasmid pTRPMTG-00.

Fig. 5 shows a construction scheme of plasmid pUCN216D.

Fig. 6 shows a construction scheme of MTG expression plasmid pUCTRPMGTG(+)D2.

Fig. 7 shows that GAT corresponding to Aspartic acid residue is deleted.

Fig. 8 shows that N-terminal amino acid is serine.

Detailed Description of the Preferred Embodiments

The proteins having a transglutaminase activity according to the present invention comprise a sequence ranging from serine residue at the second position to proline residue at the 331st position in an amino acid sequence represented by SEQ ID No. 1 as an essential sequence but the protein may further have an amino acid or amino acids after proline residue at the 331st position. Among these, the preferred is a protein consisting of an amino acid sequence of from serine residue at the second position to proline residue at the 331st position in an amino acid sequence of SEQ ID No. 1.

In these amino acid sequences, the present invention includes amino acid sequences wherein an amino acid or some amino acids are deleted, substituted or inserted as far as such amino acid sequences have a transglutaminase activity.

The DNA of the present invention encodes the above-mentioned proteins. Among these, the preferred is a DNA wherein a base sequence encoding for Arg at the forth position from the N-terminal amino acid is CGT or CGC, and a base sequence encoding for Val at the fifth position from the N-terminal amino acid is GTT or GTA. Furthermore, the preferred is a DNA wherein a base sequence encoding for the N-terminal amino acid to fifth amino acid, Ser-Asp-Asp-Arg-Val, has the following sequence.

Ser : TCT or TCC

Asp : GAC or GAT

Asp : GAC or GAT

Arg : CGT or CGC

Val : GTT or GTA

In this case, the preferred is a DNA wherein a base sequence encoding for amino acid sequence of from the N-terminal amino acid to fifth amino acid, Ser-Asp-Asp-Arg-Val, has the sequence TCT-GAC-GAT-CGT-GTT.

Furthermore, the preferred is a DNA wherein a base sequence encoding for amino acid sequence of from sixth amino acid to ninth amino acid from the N-terminal amino acid, Thr-Pro-Pro-Ala, has the following sequence.

Thr : ACT or ACC

Pro : CCA or CCG

Pro : CCA or CCG

Ala : GCT or GCA

Furthermore, the preferred is a DNA comprising a sequence ranging from thymine base at the fourth position to guanine base at the 993rd position in the base sequence of SEQ ID No. 2. In this case, more preferred is a DNA consisting of a sequence ranging from thymine base at the fourth position to guanine base at the 993rd position in the base sequence of SEQ ID No. 2.

In the DNA sequences mentioned above, nucleic acids encoding an amino acid or some amino acids may be deleted, substituted or inserted as far as such DNA encodes an amino acid sequence having a transglutaminase activity.

The recombinant DNA of the present invention has one of DNA mentioned above. In this case, the preferred is a DNA having a promoter selected from the group consisting of trp, tac, lac, trc, λ PL and T7.

The transformants of the present invention are obtained by the transformation with the above-mentioned recombinant DNA. Among these, it is preferable that a transformation be conducted by use of a multi-copy vector, and that the transformants belong to Escherichia coli.

The process for producing a protein having a transglutaminase activity according to the present invention comprises the steps of culturing one of the above-mentioned transformants in a medium to produce the protein having a transglutaminase activity and recovering the protein.

The detailed description will be further made on the present invention.

(1) It is known that the expression of MTG in the cells of a microorganism is fatal. It is also known that in the high expression of the protein in a microorganism such as E. coli, the expressed protein is inclined to be in the form of inert insoluble protein inclusion bodies. Under these circumstances, the inventors made investigations for the purpose of obtaining a high expression of MTG as an inert, insoluble protein in E. coli.

A structural gene of MTG used for achieving the high expression is a DNA containing a sequence ranging from thymine base at the fourth position to guanine base at the 993rd position in the base sequence of SEQ ID No. 2. Taking the degeneration of the genetic codon, the third letter in the degenerate codon in a domain which codes for the N-

terminal portion is converted to a codon rich in adenine and uracil and the remaining portion is comprised of a codon frequently used for E. coli in order to inhibit the formation of high-order structure of mRNA, though a DNA which codes for proteins having the same amino acid sequence can have various base sequences.

A strong promoter usually used for the production of foreign proteins is used for the expression of MTG structural gene, and a terminator is inserted into the downstream of MTG structural gene. For example, the promoters are trp, tac, lac, trc, λ PL and T7, and the terminators are trpA, lpp and T4.

For the efficient translation, the variety and number in the SD sequence, and the base composition, sequence and length in the domain between the SD sequence and initiation codon were optimized for the expression of MTG.

The domain ranging from the promoter to the terminator necessitated for the expression of MTG can be produced by a well-known chemical synthesis method. An example of the base sequence is shown in SEQ ID No. 3. In the amino acid sequence of sequence No. 3, aspartic acid residue follows the initiation codon. However, this aspartic acid residue is preferably removed as will be described below.

The present invention also provides a recombinant DNA usable for the expression of MTG.

The recombinant DNA can be produced by inserting a DNA containing the structural gene of the above-described MTG in a known expression vector selected depending on a desired expression system. The expression vector used herein is preferably a multi-copy vector.

Known expression vectors usable for the production of the recombinant DNA of the present invention include pUC19 and pHS299. An example of the recombinant DNA of the present invention obtained by integrating DNA of the present invention into pUC19 is pUCTRPMTG-02(+).

5 The present invention also relates to various transformants obtained by the introduction of the recombinant DNA.

The cells capable of forming the transformant include *E. coli* and the like.

10 An example of *E. coli* is the strain JM109 (recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, Δ (lac-proAB)/F' [traD36, proAB+, lacIq, lacZ Δ M15]).

A protein having a transglutaminase activity is produced by culturing the transformant such as that obtained by transforming *E. coli* JM109 with pUCTRPMTG-02(+) which is a vector of the present invention.

15 Examples of the medium used for the production include 2xYT medium used in the Example given below and medium usually used for culturing *E. coli* such as LB medium and M9-Casamino acid medium.

20 The culture conditions and production-inducing conditions are suitably selected depending on the kinds of the vector, promoter, host and the like. For example, for the production of a recombinant product with trp promoter, a chemical such as 3- β -indoleacrylic acid may be used for efficiently working the promoter. If necessary, glucose, Casamino acid or the like can be added in the course of the culture. Further, a chemical (ampicillin) resistant to genes which are resistant
25 to chemicals kept in plasmid can also be added in order to selectively proliferate a recombinant *E. coli*.

The protein having a transglutaminase activity, which is produced by the above-described process, is extracted from the cultured strain as follows: After the completion of the culture, the cells are collected and suspended in a buffer solution. After the treatment with lysozyme, freezing/melting, ultrasonic disintegration, etc., the thus-
5 obtained suspension of the disintegrated cells is centrifuged to divide it into a supernatant liquid and precipitates.

The protein having a transglutaminase activity is obtained in the form of a protein inclusion body and contained in the precipitates. This protein is solubilized with a denaturant or the like, the denaturant is removed and the protein is separated and purified. Examples of the denaturants usable for solubilizing the protein inclusion body produced as described above include urea (such as 8M) and guanidine hydrochloride (such as 6 M). After removing the denaturant by the dialysis or the like, the protein having a transglutaminase activity is regenerated. Solutions used for the dialysis are a phosphoric acid buffer solution, tris hydrochloride buffer solution, etc. The denaturant can be removed not only by the dialysis but also dilution, ultrafiltration or the like. The regeneration of the
10
15
20 activity is expectable by any of these techniques.

After the regeneration of the activity, the active protein can be separated and purified by a suitable combination of well-known separation and precipitation methods such as salting out, dialysis, ultrafiltration, gel filtration, SDS-polyacrylamide electrophoresis, ion exchange chromatography, affinity chromatography, reversed-phase
25 high-performance liquid chromatography and isoelectric point

electrophoresis.

(2) The present invention provides a protein having a transglutaminase activity, which has a sequence ranging from serine residue at the second position to proline residue at the 331st position in the amino acid sequence represented in SEQ ID No. 1.

The N-terminals of MTG produced by the product transformed with recombinant DNA having a DNA represented in SEQ ID No. 3 was analyzed to find that most of them contained (formyl)methionine residue of the initiation codon.

However, when a gene which encodes for an exogenous protein is expressed in E. coli, the gene is designed so that the intended protein is positioned after the methionine residue encoded by ATG which is the translation initiation signal for the gene. It is already known that N-terminal methionine residues of a natural protein obtained by the translation from genes are more efficiently cut by methionine aminopeptidase. However, the N-terminal methionine residues are not always cut in the exogenous protein.

It is known that the substrate specificity of methionine aminopeptidase varies depending on the variety of the amino acid residue positioned next to the methionine residue. When the amino acid residue positioned next to the methionine residue is alanine residue, glycine residue, serine residue or the like, the methionine residue is easily cleaved, and when the former is aspartic acid, asparagine, lysine, arginine, leucine or the like, the latter is difficultly cleaved [Nature 326, 315(1987)].

The N-terminal amino acid residue of MTG is aspartic acid

residue. When a methionine residue derived from the initiation codon is positioned directly before the aspartic acid residue, methionine aminopeptidase difficultly acts on the obtained sequence, and the N-terminal methionine residue is usually not removed but remains. However, since serine residue is arranged next to N-terminal aspartic acid in MTG, the sequence can be so designed that the amino acid residue positioned next to methionine residue derived from the initiation codon will be serine residue (an amino acid residue on which methionine aminopeptidase easily acts) by deleting aspartic acid residue. Thus, a protein having a high transglutaminase activity, from which the N-terminal methionine residue has been cleaved, can be efficiently produced.

The recombinant protein thus obtained is shorter than natural MTG by one amino acid residue, but the function of this protein is the same as that of the natural MTG. Namely, MTG activity is not lost by the lack of one amino acid. Although there is a possibility that a protein having a transglutaminase activity, from which the methionine residue has not been cleaved, gains a new antigenicity, it is generally understood that the sequence shortened by several residues does not gain a new antigenicity which natural MTG does not have. Thus, there is no problem of the safety.

In fact, a sequence of Met-Ser-Asp-Asp-Arg- was designed by deleting N-terminal aspartic acid residue from transglutaminase derived from microorganism (MTG), and this was produced in E. coli. As a result, methionine residue was efficiently removed and thereby there was obtained a protein having a sequence of Ser-Asp-

Asp-Arg-. It was confirmed that the specific activity of the thus-obtained protein is not different from that of natural MTG.

A process for producing a protein having a transglutaminase activity, which has a sequence ranging from serine residue at the second position to proline residue at the 331st position in the amino acid sequence represented in SEQ ID No. 1 will be described below.

That is, a DNA which encodes for a protein having a transglutaminase activity and having a sequence ranging from serine residue at the second position to proline residue at the 331st position in the amino acid sequence represented in SEQ ID No. 1 is employed as the MTG structural gene present on recombinant DNA usable for the expression of MTG. Concretely, a DNA having a sequence ranging from thymine base at the fourth position to guanine base at the 993rd position in the base sequence of SEQ ID No. 2 is employed.

The N-terminal sequence can be altered by an ordinary DNA recombination technique, or specific site directional mutagenesis technique, a technique wherein PCR is used for the whole or partial length of MTG gene, or a technique wherein the part of the sequence to be altered is exchanged with a synthetic DNA fragment by a restriction enzyme treatment.

The transformant thus transformed with the recombinant DNA is cultured in a medium to produce a protein having a transglutaminase activity, and the protein is recovered. The methods for the preparation of the transformant and for the production of the protein are the same as those described above.

Since the protein thus produced has a sequence of Met-Ser-. . .

from which the methionine residue is easily cleaved with methionine amino-peptidase, the methionine residue is cleaved in the cell of E. coli to obtain a protein that starts with serine residue.

Although MTG having N-terminal methionine residue is not present in the nature, the inventors have found that in some of natural MTG, aspartic acid residue is deleted to have N-terminal serine. Although a protein having N-terminal methionine residue is thus different from natural MTG in the sequence, a protein having N-terminal serine residue is included in the sequences of natural MTG and, in addition, a protein having such a sequence is actually present in the nature. Thus, it can be said that such MTG is equal to natural MTG. Namely, in the production of an enzyme to be used for foods, such as MTG, in which protein antigenicity is a serious problem, it is important to produce a protein having transglutaminase activity and also having a sequence equal to that of natural MTG, or in other words, to produce a sequence from which the N-terminal methionine residue was cleaved.

The following Examples will further illustrate the present invention, which by no means limit the invention.

Example

Mass production of MTG in E. coli:

<1> Construction of MTG expression plasmid PTRPMTG-01:

MTG gene has been already completely synthesized, taking the frequency of using codons of E. coli and yeast into consideration (J. P. KOKAI No. Hei 5-199883). However, the gene sequence thereof was not optimum for the expression in E. coli. Namely, all of codons of thirty arginine residues were AGA (minor codons). Under these conditions,

about 200 bases from the N-terminal of MTG gene were resynthesized to become a sequence optimum for the expression of *E. coli*.

As a promoter for transcribing MTG gene, trp promoter capable of easily deriving the transcription in a medium lacking tryptophane was used. Plasmid pTTG2-22 (J. P. KOKAI No. Hei 6-225775) for the high expression of transglutaminase (TG) gene of *Pagrus major* was obtained with trp promoter. The sequence in the upstream of the TG gene of *Pagrus major* was designed so that a foreign protein is highly expressed in *E. coli*.

In the construction of pTRPMTG-01, the DNA fragment from ClaI site in the downstream of trp promoter to BglII site in the downstream of *Pagrus major*'s TG expression plasmid pTTG2-22 (J. P. KOKAI Hei 6-225775) was replaced with the ClaI/HpaI fragment of the synthetic DNA gene and the HpaI/BamHI fragment (small) of pGEM15BTG (J. P. KOKAI Hei 6-30771).

The ClaI/HpaI fragment of the Synthetic DNA gene has a base sequence from ClaI site in the downstream of trp promoter of pTTG2-22 to translation initiation codon, and 216 bases from the N-terminal of MTG gene. The base sequence in MTG structural gene was determined with reference to the frequency of using codon in *E. coli* so as to be optimum for the expression in *E. coli*. However, for avoiding the high-order structure of mRNA, the third letter of the degenerated codon in the domain of encoding the N-terminal part was converted to a codon rich in adenine and uracil so as to avoid the arrangement of the same bases as far as possible.

The ClaI/HpaI fragment of the Synthetic DNA gene was so designed

that it had EcoRI and HindIII sites at the terminal. The designed gene was divided into blocks each comprising about 40 to 50 bases so that the + chain and the - chain overlapped each other. Twelve DNA fragments corresponding to each sequence were synthesized (SEQ ID Nos. 4 to 15).

5 5' terminal of the synthetic DNA was phosphatized. Synthetic DNA fragments to be paired therewith were annealed, and they were connected with each other. After the acrylamide gel electrophoresis, the DNA fragments of an intended size was taken out and integrated in EcoRI/HindIII sites of pUC19. The sequence was confirmed and the correct one was named pUCN216. From the pUCN216, a ClaI/HpaI fragment (small) was taken out and used for the construction of pTRPMTG-01.

10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340
9345
9350
9355
9360
9365
9370
9375
9380
9385
9390
9395
9400
9405
9410
9415
9420
9425
9430
9435
9440
9445
9450
9455
9460
9465
9470
9475
9480
9485
9490
9495
9500
9505
9510
9515
9520
9525
9530
9535
9540
9545
9550
9555
9560
9565
9570
9575
9580
9585
9590
9595
9600
9605
9610
9615
9620
9625
9630
9635
9640
9645
9650
9655
9660
9665
9670
9675
9680
9685
9690
9695
9700
9705
9710
9715
9720
9725
9730
9735
9740
9745
9750
9755
9760
9765
9770
9775
9780
9785
9790
9795
9800
9805
9810
9815
9820
9825
9830
9835
9840
9845
9850
9855
9860
9865
9870
9875
9880
9885
9890
9895
9900
9905
9910
9915
9920
9925
9930
9935
9940
9945
9950
9955
9960
9965
9970
9975
9980
9985
9990
9995
10000
10005
10010
10015
10020
10025
10030
10035
10040
10045
10050
10055
10060
10065
10070
10075
10080
10085
10090
10095
10100
10105
10110
10115
10120
10125
10130
10135
10140
10145
10150
10155
10160
10165
10170
10175
10180
10185
10190
10195
10200
10205
10210
10215
10220
10225
10230

acrylamide gel electrophoresis, DNA of an intended size was taken out and integrated in EcoRI/HindIII sites of pUC19. The base sequence of each of them was confirmed and the correct ones were named pUCB1, B2, B3 and B4. As shown in Fig. 2, B1 was connected with B2, and B3 was connected with B4. By replacing a corresponding part of pTRPMTG-01 therewith, pTRPMTG-02 was constructed. The sequence of the high expression MTG gene present on pTRPMTG-02 is shown in SEQ ID No. 3.

<3> Construction of MTG expression plasmid pUCTRPMTG-02(+), (-):

Since E. coli JM109 which keeps the pTRPMTG-02 also did not highly express MTG, the plasmid was multi-copied. Eco109I fragment (small) containing trp promoter of pTRPMTG-02 was smoothened and then integrated into HincII site of pUC19 which is a multi-copy plasmid. pUCTRPMTG-02(+) in which lacZ promoter and trp promoter were in the same direction, and pUCTRPMTG-02(-) in which they were in the opposite direction to each other were constructed.

<4> Expression of MTG:

E. coli JM109 transformed with pUCTRPMTG-02(+) and pUC19 was cultured by shaking in 3 ml of 2xYT medium containing 150 μ g/ml of ampicillin at 37°C for ten hours (pre-culture). 0.5 ml of the culture suspension was added to 50 ml of 2xYT medium containing 150 μ g/ml of ampicillin, and the shaking culture was conducted at 37°C for 20 hours.

The cells were collected from the culture suspension and broken by ultrasonic disintegration. The results of SDS-polyacrylamide electrophoresis of the whole fraction, and supernatant and precipitation fractions both obtained by the centrifugation are shown in Fig. 3. The high expression of the protein having a molecular weight

equal to that of MTG was recognized in the whole fraction of broken
pUCTRPMTG-02(+)/JM109 cells and the precipitate fraction obtained by
the centrifugation. It was confirmed by the western blotting that the
protein was reactive with mice anti-MTG antibody. The expression of
the protein was 500 to 600 mg/L. A sufficient, high expression was
obtained even when 3- β -indole acrylic acid was not added to the
production medium.

Further, the western blotting was conducted with MTG antibody
against mouse to find that MTG was expressed only slightly in the
supernatant fraction obtained by the centrifugation and that the
expressed MTG was substantially all in the form of insoluble protein
inclusion bodies.

<5> Construction of MTG expression plasmid pTRPMTG-00:

To prove that the change in codon of MTG gene caused a
remarkable increase in the expression, pTRPMTG-00 corresponding to
pTRPMTG-02 but in which MTG gene was changed to a gene sequence
completely synthesized before (J. P. KOKAI No. Hei 6-30771) was
constructed.

pTRPMTG-00 was constructed by connecting PvuII/PstI fragment
(small) from Pagrus major's TG expression plasmid pTRPMTG-02 with
PstI/HindIII fragment (small, including PvuII site) and PvuII/HindIII
fragment (small) of pGEM15BTG (J. P. KOKAI No. Hei 6-30771).

<6> Construction of MTG expression plasmid pUCTRPMTG-00(+), (-):

pTRPMTG-00 was multi-copied. EcoO109I fragment (small)
containing trp promoter and trpA terminator of pTRPMTG-00 was smoothened
and then integrated into HincII site of pUC19 which is a multi-copy

plasmid. pUCTRPMTG-00(+) in which lacZ promoter and trp promoter were in the same direction, and pUCTRPMTG-00(-) in which they were in the opposite direction to each other were constructed.

<7> Comparison of MTG expressions:

5 E. coli JM109 transformed with pUCTRPMTG-02 (+) or (-), pUCTRPMTG-00 (+) or (-), pTRPMTG-02, pTRPMTG-01, pTRPMTG-00 or pUC19 was cultured by shaking in 3 ml of 2xYT medium containing 150 μ g/ml of ampicillin at 37°C for ten hours (pre-culture). 0.5 ml of the culture suspension was added to 50 ml of 2xYT medium containing 150 μ g/ml of ampicillin, and the shaking culture was conducted at 37 °C for 20 hours.

10 The cells were collected from the culture suspension, and MTG expression thereof was determined to obtain the results shown in Table 1. It was found that the newly constructed E. coli containing pTRPMTG-00, pUCTRPMTG-00 (+) or (-) did not highly express MTG. This result indicate that it is necessary for the high expression of MTG to change the codon of MTG gene into a codon for E. coli and also to multi-copy the plasmid.

Table 1

Strain	MTG expression
pUCTRPMTG-02(+)/JM109	+ + +
pUCTRPMTG-02(-)/JM109	+ + +
pUCTRPMTG-00(+)/JM109	+
pUCTRPMTG-00(-)/JM109	+
pTRPMTG-02/JM109	+
pTRPMTG-01/JM109	+
pTRPMTG-00/JM109	-
pUC19/JM109	-

+ + + : at least 300 mg/l

+ : 5 mg/l or below

- : no expression

<8> Analysis of N-terminal amino acid of expressed MTG:

The N-terminal amino acid residue of the protein inclusion bodies of expressed MTG was analyzed to find that about 60 % of the sequence of N-terminal was methionine residue and about 40 % thereof was formylmethionine residue. (Formyl)methionine residue corresponding to the initiation codon was removed by a technical idea described below.

<9> Deletion of N-terminal aspartic acid residue of MTG:

A base sequence corresponding to aspartic acid residue (the N-terminal of MTG) was deleted by PCR using pUCN216 containing 216 bases as the template. pUCN216 is a plasmid obtained by cloning about 216

bp's containing ClaI-HpaI fragment of N-terminal of MTG in EcoRI/HindIII site of pUC19. pF01 (SEQ ID No. 56) and pR01 (SEQ ID No. 57) are primers each having a sequence in the vector. pDELD (SEQ ID No. 58) is that obtained by deleting a base sequence corresponding to Asp residue. pHD01 (SEQ ID No. 59) is that obtained by replacing C with G not to include HindIII site. pF01 and pDELD are sense primers and pR01 and pHD01 are antisense primers.

35 cycles of PCR of a combination of pF01 and pHD01, and a combination of pELD and pR01 for pUCN216 was conducted at 94 °C for 30 seconds, at 55°C for one minute and at 72 °C for two minutes. Each PCR product was extracted with phenol/chloroform, precipitated with ethanol and dissolved in 100 μ l of H₂O.

1 μ l of each of the PCR products was taken, and they were mixed together. After the heat denaturation at 94 °C for 10 minutes, 35 cycles of PCR of a combination of pF01 and pHD01 was conducted at 94°C for 30 seconds, at 55°C for one minute and at 72 °C for two minutes.

The second PCR product was extracted with phenol/chloroform, precipitated with ethanol, and treated with HindIII and EcoRI. After pUC19 subcloning, pUCN216D was obtained (Fig. 5). The sequence of the obtained pUCN216D was confirmed to be the intended one.

<10> Construction of the plasmid encoding for MTG which lacks N-terminal aspartic acid:

Eco0109I/HpaI fragment (small) of pUCN216D was combined with Eco0109I/HpaI fragment (large) of pUCB1-1 (plasmid obtained by cloning HpaII/BglIII fragment of MTG gene in EcoRI/HindIII site of pUC19) to obtain pUCNB1-2D. Further, ClaI/BglIII fragment (small) of pUCNB1-2D was

combined with ClaI/BglIII fragment (large) of pUCTRPMTG-02(+) which is a plasmid of high MTG expression to obtain pUC TRPMTG(+)D2, the expression plasmid of MTG which lacks N-terminal aspartic acid(Fig. 6). As a result, a plasmid containing MTG gene lacking GAI corresponding to aspartic acid residue as shown in Fig. 7 was obtained.

<11> Expression of the plasmid encoding for MTG which lacks N-terminal aspartic acid:

E. coli JM109 transformed with pUCTRPMTG(+)D2 was cultured by shaking in 3 ml of 2xYT medium containing $150 \mu\text{g/ml}$ of ampicillin at 37°C for ten hours (pre-culture). 0.5 ml of the culture suspension was added to 50 ml of 2xYT medium containing $150 \mu\text{g/ml}$ of ampicillin, and the shaking culture was conducted at 37°C for 20 hours. The cells were broken by the ultrasonic disintegration. The results of the dyeing with Coomassie Brilliant Blue dyeing and Western blotting with mouse antiMTG antibody of the thus obtained supernatant liquid and precipitate indicated that MTG protein lacking N-terminal aspartic acid residue was detected in the precipitate obtained by the ultrasonic disintegration, namely in the insoluble fraction. This fact suggests that MTG protein lacking N-terminal aspartic acid residue was accumulated as protein inclusion bodies in the cells.

The N-terminal amino acid sequence of the protein inclusion bodies was analyzed to find that about 90 % thereof was serine as shown in Fig.8.

The results of the analysis of N-terminal amino acids of expressed MTG obtained in <8> and <11> were compared with each other as shown in Table 2. It was found that by deleting the N-terminal

aspartic acid residue from MTG, the initiation methionine added to the N-terminal of the expressed MTG was efficiently removed.

Table 2

Strain	N-terminal amino acid			
	f-Met	Met	Asp	Ser
pUCTRPMTG-02(+)/JM109	40 %	60 %	N.D.	
pUCTRPMTG(+)/D2/JM109	N.D.	10 %	-	90 %

<12> Solubilization of MTG inclusion bodies lacking N-terminal aspartic acid residue, renaturation of activity and determination of specific activity:

MTG inclusion bodies lacking aspartic acid was partially purified by repeating the centrifugation several times, and then dissolved in 8 M urea [50 mM phosphate buffer (pH 5.5)] to obtain the 2 mg/ml solution. Precipitates were removed from the solution by the centrifugation and the solution was diluted to a concentration of 0.5 M urea with 50 mM phosphate buffer (pH 5.5). The diluted solution was further dialyzed with 50 mM phosphate buffer (pH 5.5) to remove urea. According to Mono S column test, the peak having TG activity was eluted when NaCl concentration was in the range of 100 to 150 mM. The specific activity of the fraction was determined by the hydroxamate method to find that the specific activity of the aspartic acid residue-lacking MTG was about 30 U/mg. This is equal to the specific activity of natural MTG. It is thus apparent that the lack of aspartic acid residue exerts no influence on the specific activity.

SEQUENCE LISTING

INFORMATION FOR SEQ ID NO:1:

SEQUENCE CHARACTERISTICS:

LENGTH:331

TYPE:amino acid

TOPOLOGY:linear

MOLECULE TYPE:peptide

SEQUENCE DESCRIPTION: SEQ ID NO:1

Asp Ser Asp Asp Arg Val Thr Pro Pro Ala Glu Pro Leu Asp Arg Met

1 5 10 15

Pro Asp Pro Tyr Arg Pro Ser Tyr Gly Arg Ala Glu Thr Val Val Asn

20 25 30

Asn Tyr Ile Arg Lys Trp Gln Gln Val Tyr Ser His Arg Asp Gly Arg

35 40 45

Lys Gln Gln Met Thr Glu Glu Gln Arg Glu Trp Leu Ser Tyr Gly Cys

50 55 60

Val Gly Val Thr Trp Val Asn Ser Gly Gln Tyr Pro Thr Asn Arg Leu

65 70 75 80

Ala Phe Ala Ser Phe Asp Glu Asp Arg Phe Lys Asn Glu Leu Lys Asn

85 90 95

Gly Arg Pro Arg Ser Gly Glu Thr Arg Ala Glu Phe Glu Gly Arg Val

100

105

110

Ala Lys Glu Ser Phe Asp Glu Glu Lys Gly Phe Gln Arg Ala Arg Glu

115

120

125

Val Ala Ser Val Met Asn Arg Ala Leu Glu Asn Ala His Asp Glu Ser

130

135

140

Ala Tyr Leu Asp Asn Leu Lys Lys Glu Leu Ala Asn Gly Asn Asp Ala

145

150

155

160

Leu Arg Asn Glu Asp Ala Arg Ser Pro Phe Tyr Ser Ala Leu Arg Asn

165

170

175

Thr Pro Ser Phe Lys Glu Arg Asn Gly Gly Asn His Asp Pro Ser Arg

180

185

190

Met Lys Ala Val Ile Tyr Ser Lys His Phe Trp Ser Gly Gln Asp Arg

195

200

205

Ser Ser Ser Ala Asp Lys Arg Lys Tyr Gly Asp Pro Asp Ala Phe Arg

210

215

220

Pro Ala Pro Gly Thr Gly Leu Val Asp Met Ser Arg Asp Arg Asn Ile

225

230

235

240

Pro Arg Ser Pro Thr Ser Pro Gly Glu Gly Phe Val Asn Phe Asp Tyr

245

250

255

Gly Trp Phe Gly Ala Gln Thr Glu Ala Asp Ala Asp Lys Thr Val Trp
 260 265 270

Thr His Gly Asn His Tyr His Ala Pro Asn Gly Ser Leu Gly Ala Met
 275 280 285

His Val Tyr Glu Ser Lys Phe Arg Asn Trp Ser Glu Gly Tyr Ser Asp
 290 295 300

Phe Asp Arg Gly Ala Tyr Val Ile Thr Phe Ile Pro Lys Ser Trp Asn
 305 310 315 320

Thr Ala Pro Asp Lys Val Lys Gln Gly Trp Pro
 325 330

INFORMATION FOR SEQ ID NO:2:

SEQUENCE CHARACTERISTICS:

LENGTH: 993

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

FEATURE

FEATURE KEY: CDS

LOCATION: 1..993

IDENTIFICATION METHOD: S

SEQUENCE DESCRIPTION: SEQ ID NO:2

GAT TCT GAC GAT CGT GTT ACT CCA CCA GCT GAA CCA CTG GAT CGT ATG

48

Asp Ser Asp Asp Arg Val Thr Pro Pro Ala Glu Pro Leu Asp Arg Met

1

5

10

15

CCA GAT CCA TAT CGT CCA TCT TAT GGT CGT GCT GAA ACT GTT GTT AAT

96

Pro Asp Pro Tyr Arg Pro Ser Tyr Gly Arg Ala Glu Thr Val Val Asn

20

25

30

AAT TAT ATT CGT AAA TGG CAA CAA GTT TAT TCT CAT CGT GAT GGT CGT

144

Asn Tyr Ile Arg Lys Trp Gln Gln Val Tyr Ser His Arg Asp Gly Arg

35

40

45

AAA CAA CAA ATG ACT GAA GAA CAA CGT GAA TGG CTG TCT TAT GGT TGC

192

Lys Gln Gln Met Thr Glu Glu Gln Arg Glu Trp Leu Ser Tyr Gly Cys

50

55

60

GTT GGT GTT ACT TGG GTT AAC TCT GGT CAG TAT CCG ACT AAC CGT CTG

240

Val Gly Val Thr Trp Val Asn Ser Gly Gln Tyr Pro Thr Asn Arg Leu

65

70

75

80

GCA TTC GCT TCC TTC GAT GAA GAT CGT TTC AAG AAC GAA CTG AAG AAC

288

Ala Phe Ala Ser Phe Asp Glu Asp Arg Phe Lys Asn Glu Leu Lys Asn

85

90

95

GGT CGT CCG CGT TCT GGT GAA ACT CGT GCT GAA TTC GAA GGT CGT GTT

336

Gly Arg Pro Arg Ser Gly Glu Thr Arg Ala Glu Phe Glu Gly Arg Val

100

105

110

GCT AAG GAA TCC TTC GAT GAA GAG AAA GGC TTC CAG CGT GCT CGT GAA

384

Ala Lys Glu Ser Phe Asp Glu Glu Lys Gly Phe Gln Arg Ala Arg Glu

115

120

125

GTT GCT TCT GTT ATG AAC CGT GCT CTA GAG AAC GCT CAT GAT GAA TCT

432

Val Ala Ser Val Met Asn Arg Ala Leu Glu Asn Ala His Asp Glu Ser

130

135

140

GCT TAC CTG GAT AAC CTG AAG AAG GAA CTG GCT AAC GGT AAC GAT GCT

480

Ala Tyr Leu Asp Asn Leu Lys Lys Glu Leu Ala Asn Gly Asn Asp Ala

145

150

155

160

CTG CGT AAC GAA GAT GCT CGT TCT CCG TTC TAC TCT GCT CTG CGT AAC

528

Leu Arg Asn Glu Asp Ala Arg Ser Pro Phe Tyr Ser Ala Leu Arg Asn

165

170

175

ACT CCG TCC TTC AAA GAA CGT AAC GGT GGT AAC CAT GAT CCG TCT CGT

576

Thr Pro Ser Phe Lys Glu Arg Asn Gly Gly Asn His Asp Pro Ser Arg

180

185

190

ATG AAA GCT GTT ATC TAC TCT AAA CAT TTC TGG TCT GGT CAG GAT AGA

624

Met Lys Ala Val Ile Tyr Ser Lys His Phe Trp Ser Gly Gln Asp Arg

195

200

205

TCT TCT TCT GCT GAT AAA CGT AAA TAC GGT GAT CCG GAT GCA TTC CGT

672

Ser Ser Ser Ala Asp Lys Arg Lys Tyr Gly Asp Pro Asp Ala Phe Arg

210

215

220

CCG GCT CCG GGT ACT GGT CTG GTA GAC ATG TCT CGT GAT CGT AAC ATC

720

Pro Ala Pro Gly Thr Gly Leu Val Asp Met Ser Arg Asp Arg Asn Ile

225

230

235

240

CCG CGT TCT CCG ACT TCT CCG GGT GAA GGC TTC GTT AAC TTC GAT TAC

768

Pro Arg Ser Pro Thr Ser Pro Gly Glu Gly Phe Val Asn Phe Asp Tyr

245

250

255

GGT TGG TTC GGT GCT CAG ACT GAA GCT GAT GCT GAT AAG ACT GTA TGG

816

Gly Trp Phe Gly Ala Gln Thr Glu Ala Asp Ala Asp Lys Thr Val Trp

260

265

270

ACC CAT GGT AAC CAT TAC CAT GCT CCG AAC GGT TCT CTG GGT GCT ATG

864

Thr His Gly Asn His Tyr His Ala Pro Asn Gly Ser Leu Gly Ala Met

275

280

285

CAT GTA TAC GAA TCT AAA TTC CGT AAC TGG TCT GAA GGT TAC TCT GAC

912

His Val Tyr Glu Ser Lys Phe Arg Asn Trp Ser Glu Gly Tyr Ser Asp

290

295

300

TTC GAT CGT GGT GCT TAC GTT ATC ACC TTC ATT CCG AAA TCT TGG AAC

960

Phe Asp Arg Gly Ala Tyr Val Ile Thr Phe Ile Pro Lys Ser Trp Asn

305

310

315

320

ACT GCT CCG GAC AAA GTT AAA CAG GGT TGG CCG

993

Thr Ala Pro Asp Lys Val Lys Gln Gly Trp Pro

325

330

INFORMATION FOR SEQ ID NO:3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1518

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

FEATURE

FEATURE KEY: CDS

LOCATION: 87..1082

IDENTIFICATION METHOD:S

SEQUENCE DESCRIPTION: SEQ ID NO:3

TTCCCCTGTT GACAATTAAT CATCGAACTA GTTAACTAGT ACGCAAGTTC ACGTAAAAAG

60

GGTATCGATT AGTAAGGAGG TTTAAA ATG GAT TCT GAC GAT CGT GTT ACT CCA

113

Met Asp Ser Asp Asp Arg Val Thr Pro

1

5

CCA GCT GAA CCA CTG GAT CGT ATG CCA GAT CCA TAT CGT CCA TCT TAT

161

Pro Ala Glu Pro Leu Asp Arg Met Pro Asp Pro Tyr Arg Pro Ser Tyr

10

15

20

25

GGT CGT GCT GAA ACT GTT GTT AAT AAT TAT ATT CGT AAA TGG CAA CAA

209

Gly Arg Ala Glu Thr Val Val Asn Asn Tyr Ile Arg Lys Trp Gln Gln

30

35

40

GTT TAT TCT CAT CGT GAT GGT CGT AAA CAA CAA ATG ACT GAA GAA CAA

257

Val Tyr Ser His Arg Asp Gly Arg Lys Gln Gln Met Thr Glu Glu Gln

45

50

55

CGT GAA TGG CTG TCT TAT GGT TGC GTT GGT GTT ACT TGG GTT AAC TCT

305

Arg Glu Trp Leu Ser Tyr Gly Cys Val Gly Val Thr Trp Val Asn Ser

60

65

70

GGT CAG TAT CCG ACT AAC CGT CTG GCA TTC GCT TCC TTC GAT GAA GAT

353

Gly Gln Tyr Pro Thr Asn Arg Leu Ala Phe Ala Ser Phe Asp Glu Asp

75

80

85

CGT TTC AAG AAC GAA CTG AAG AAC GGT CGT CCG CGT TCT GGT GAA ACT

401

Arg Phe Lys Asn Glu Leu Lys Asn Gly Arg Pro Arg Ser Gly Glu Thr

90

95

100

105

CGT GCT GAA TTC GAA GGT CGT GTT GCT AAG GAA TCC TTC GAT GAA GAG

449

Arg Ala Glu Phe Glu Gly Arg Val Ala Lys Glu Ser Phe Asp Glu Glu

110

115

120

AAA GGC TTC CAG CGT GCT CGT GAA GTT GCT TCT GTT ATG AAC CGT GCT

497

Lys Gly Phe Gln Arg Ala Arg Glu Val Ala Ser Val Met Asn Arg Ala

125

130

135

CTA GAG AAC GCT CAT GAT GAA TCT GCT TAC CTG GAT AAC CTG AAG AAG

545

Leu Glu Asn Ala His Asp Glu Ser Ala Tyr Leu Asp Asn Leu Lys Lys

140

145

150

GAA CTG GCT AAC GGT AAC GAT GCT CTG CGT AAC GAA GAT GCT CGT TCT

593

Glu Leu Ala Asn Gly Asn Asp Ala Leu Arg Asn Glu Asp Ala Arg Ser

155

160

165

CCG TTC TAC TCT GCT CTG CGT AAC ACT CCG TCC TTC AAA GAA CGT AAC

641

Pro Phe Tyr Ser Ala Leu Arg Asn Thr Pro Ser Phe Lys Glu Arg Asn

170

175

180

185

GGT GGT AAC CAT GAT CCG TCT CGT ATG AAA GCT GTT ATC TAC TCT AAA

689

Gly Gly Asn His Asp Pro Ser Arg Met Lys Ala Val Ile Tyr Ser Lys

190

195

200

CAT TTC TGG TCT GGT CAG GAT AGA TCT TCT TCT GCT GAT AAA CGT AAA

737

His Phe Trp Ser Gly Gln Asp Arg Ser Ser Ser Ala Asp Lys Arg Lys

205

210

215

TAC GGT GAT CCG GAT GCA TTC CGT CCG GCT CCG GGT ACT GGT CTG GTA

785

Tyr Gly Asp Pro Asp Ala Phe Arg Pro Ala Pro Gly Thr Gly Leu Val

220

225

230

GAC ATG TCT CGT GAT CGT AAC ATC CCG CGT TCT CCG ACT TCT CCG GGT

833

Asp Met Ser Arg Asp Arg Asn Ile Pro Arg Ser Pro Thr Ser Pro Gly

235

240

245

GAA GGC TTC GTT AAC TTC GAT TAC GGT TGG TTC GGT GCT CAG ACT GAA

881

Glu Gly Phe Val Asn Phe Asp Tyr Gly Trp Phe Gly Ala Gln Thr Glu
 250 255 260 265

GCT GAT GCT GAT AAG ACT GTA TGG ACC CAT GGT AAC CAT TAC CAT GCT
 929

Ala Asp Ala Asp Lys Thr Val Trp Thr His Gly Asn His Tyr His Ala
 270 275 280

CCG AAC GGT TCT CTG GGT GCT ATG CAT GTA TAC GAA TCT AAA TTC CGT
 977

Pro Asn Gly Ser Leu Gly Ala Met His Val Tyr Glu Ser Lys Phe Arg
 285 290 295

AAC TGG TCT GAA GGT TAC TCT GAC TTC GAT CGT GGT GCT TAC GTT ATC
 1025

Asn Trp Ser Glu Gly Tyr Ser Asp Phe Asp Arg Gly Ala Tyr Val Ile
 300 305 310

ACC TTC ATT CCG AAA TCT TGG AAC ACT GCT CCG GAC AAA GTT AAA CAG
 1073

Thr Phe Ile Pro Lys Ser Trp Asn Thr Ala Pro Asp Lys Val Lys Gln
 315 320 325

GGT TGG CCG TAATGAAAGC TTGGATCTCT AATTACTGGA CTTCACACAG ACTAAAATAG
 1131

Gly Trp Pro
 330

ACATATCTTA TATTATGTGA TTTTGTGACA TTTCCTAGAT GTGAGGTGGA GGTGATGTAT

1191
 AAGGTAGATG ATGATCCTCT ACGCCGGACG CATCGTGGCC GGCATCACCG GCGCCACAGG
 1251
 TCGCGTTGCT GCGCCTATA TCGCCGACAT CACCGATGGG GAAGATCGGG CTCGCCACTT
 1311
 CGGGCTCATG AGCGCTTGTT TCGGCGTGGG TATGGTGGCA GGCCCCGTGG CCGGGGGACT
 1371
 GTTGGGCGCC ATCTCCTTGC ATGCACCATT CCTTGGCGCG GCGGTGCTCA ACGGCCTCAA
 1431
 CCTACTACTG GGCTGCTTCC TAATGCAGGA GTCGCATAAG GGAGAGCGTC GAGAGCCCGC
 1491
 CTAATGAGCG GGCTTTTTTT TCAGCTG
 1518

INFORMATION FOR SEQ ID NO:4:

SEQUENCE CHARACTERISTICS:

LENGTH:39

TYPE:nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:4

AATTCATCGA TTAGTAAGGA GGTTTAAAT GGATTCTGA

39

INFORMATION FOR SEQ ID NO:5:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:5

CGATCGTCAG AATCCATTTT AAACCTCCTT ACTAATCGAT G

41

INFORMATION FOR SEQ ID NO:6:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:6

CGATCGTGTT ACTCCACCAG CTGAACCACT GGATCGTATG C

41

INFORMATION FOR SEQ ID NO:7:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:7

GATCTGGCAT ACGATCCAGT GGTTCAGCTG GTGGAGTAAC A

41

INFORMATION FOR SEQ ID NO:8:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:8

CAGATCCATA TCGTCCATCT TATGGTCGTG CTGAAACTGT T

41

INFORMATION FOR SEQ ID NO:9:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:9

ATTAACAACA GTTTCAGCAC GACCATAAGA TGGACGATAT G

41

INFORMATION FOR SEQ ID NO:10:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:10

GTTAATAATT ATATTCGTAA ATGGCAACAA GTTTATTCTC A

41

INFORMATION FOR SEQ ID NO:11:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:11

TCACGATGAG AATAAACTTG TTGCCATTTA CGAATATAAT T

INFORMATION FOR SEQ ID NO:12:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:12

TCGTGATGGT CGTAAACAAC AATGACTGA AGAACAACGT G

41

INFORMATION FOR SEQ ID NO:13:

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:13

GCCATTCACG TTGTTCTTCA GTCATTGTGTT GTTACGACC A

41

INFORMATION FOR SEQ ID NO:14:

SEQUENCE CHARACTERISTICS:

LENGTH:42

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:14

AATGGCTGTC TTATGGTTGC GTTGGTGTTA CTTGGGTAA CA

42

INFORMATION FOR SEQ ID NO:15:

SEQUENCE CHARACTERISTICS:

LENGTH:40

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:15

AGCTTGTTAA CCCAAGTAAC ACCAACGCAA CCATAAGACA

40

INFORMATION FOR SEQ ID NO:16:

SEQUENCE CHARACTERISTICS:

LENGTH:38

TYPE: nucleic acid
STRANDEDNESS:single
TOPOLOGY: linear
MOLECULE TYPE:other nucleic acid synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:16

AATTCGTAA CTCTGGTCAG TATCCGACTA ACCGTCTG

38

INFORMATION FOR SEQ ID NO:17:

SEQUENCE CHARACTERISTICS:

LENGTH:41
TYPE: nucleic acid
STRANDEDNESS:single
TOPOLOGY: linear
MOLECULE TYPE:other nucleic acid synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:17

CGAATGCCAG ACGGTTAGTC GGATACTGAC CAGAGTTAAC G

41

INFORMATION FOR SEQ ID NO:18:

SEQUENCE CHARACTERISTICS:

LENGTH:49
TYPE: nucleic acid
STRANDEDNESS:single
TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:18

GCATTCGCTT CCTTCGATGA AGATCGTTTC AAGAACGAAC TGAAGAACG

49

INFORMATION FOR SEQ ID NO:19:

SEQUENCE CHARACTERISTICS:

LENGTH:49

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:19

GGACGACCGT TCTTCAGTTC GTTCTTGAAA CGATCTTCAT CGAAGGAAG

49

INFORMATION FOR SEQ ID NO:20:

SEQUENCE CHARACTERISTICS:

LENGTH:35

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:20

GTCGTCCGCG TTCTGGTGAA ACTCGTGCTG AATTC

INFORMATION FOR SEQ ID NO:21:

SEQUENCE CHARACTERISTICS:

LENGTH:35

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:21

GACCTTCGAA TTCAGCACGA GTTTCACCAG AACGC

35

INFORMATION FOR SEQ ID NO:22:

SEQUENCE CHARACTERISTICS:

LENGTH:48

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:22

GAAGGTCGTG TTGCTAAGGA ATCCTTCGAT GAAGAGAAAG GCTTCCAG

48

INFORMATION FOR SEQ ID NO:23:

SEQUENCE CHARACTERISTICS:

LENGTH:48

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:23

GAGCACGCTG GAAGCCTTC TCTTCATCGA AGGATTCCTT AGCAACAC

48

INFORMATION FOR SEQ ID NO:24:

SEQUENCE CHARACTERISTICS:

LENGTH:42

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:24

CGTGCTCGTG AAGTTGCTTC TGTATGAAC CGTGCTCTAG AA

42

INFORMATION FOR SEQ ID NO:25:

SEQUENCE CHARACTERISTICS:

LENGTH:39

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:25

AGCTTTCTAG AGCACGGTTC ATAACAGAAG CAACTTCAC

39

INFORMATION FOR SEQ ID NO:26:

SEQUENCE CHARACTERISTICS:

LENGTH:45

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:26

AATTCCTAG AGAACGCTCA TGATGAATCT GCTTACCTGG ATAAC

45

INFORMATION FOR SEQ ID NO:27:

SEQUENCE CHARACTERISTICS:

LENGTH:50

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:27

CTTCTTCAGG TTATCCAGGT AAGCAGATTC ATCATGAGCG TTCTCTAGAG

50

INFORMATION FOR SEQ ID NO:28:

SEQUENCE CHARACTERISTICS:

LENGTH:49

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:28

CTGAAGAAGG AACTGGCTAA CGTAACGAT GCTCTGCGTA ACGAAGATG

49

INFORMATION FOR SEQ ID NO:29:

SEQUENCE CHARACTERISTICS:

LENGTH:49

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:29

GAGAACGAGC ATCTTCGTTA CGCAGAGCAT CGTTACCGTT AGCCAGTTC

INFORMATION FOR SEQ ID NO:30

SEQUENCE CHARACTERISTICS:

LENGTH:40

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:30

CTCGTTCTCC GTTCTACTCT GCTCTGCGTA ACACTCCGTC

40

INFORMATION FOR SEQ ID NO:31:

SEQUENCE CHARACTERISTICS:

LENGTH:39

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:31

CTTTGAAGGA CGGAGTGTTA CGCAGAGCAG AGTAGAACG

39

INFORMATION FOR SEQ ID NO:32:

SEQUENCE CHARACTERISTICS:

LENGTH:47

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:32

CTTCAAAGAA CGTAACGGTG GTAACCATGA TCCGTCTCGT ATGAAAG

47

INFORMATION FOR SEQ ID NO:33:

SEQUENCE CHARACTERISTICS:

LENGTH:47

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:33

GATAACAGCT TTCATACGAG ACGGATCATG GTTACCACCG TTACGTT

47

INFORMATION FOR SEQ ID NO:34:

SEQUENCE CHARACTERISTICS:

LENGTH:45

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:34

CTGTTATCTA CTCTAAACAT TTCTGGTCTG GTCAGGATAG ATCTA

45

INFORMATION FOR SEQ ID NO:35:

SEQUENCE CHARACTERISTICS:

LENGTH: 41

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:35

AGCTTAGATC TATCCTGACC AGACCAGAAA TGTTTAGAGT A

41

INFORMATION FOR SEQ ID NO:36:

SEQUENCE CHARACTERISTICS:

LENGTH: 42

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:36

AATTCAGATC TTCTTCTGCT GATAAACGTA AATACGGTGA TC

42

INFORMATION FOR SEQ ID NO:37:

SEQUENCE CHARACTERISTICS:

LENGTH:44

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:37

CATCCGGATC ACCGTATTTA CGTTTATCAG CAGAAGAAGA TCTG

44

INFORMATION FOR SEQ ID NO:38:

SEQUENCE CHARACTERISTICS:

LENGTH:48

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:38

CGGATGCATT CCGTCCGGCT CCGGGTACTG GTCIGGTAGA CATGTCTC

48

INFORMATION FOR SEQ ID NO:39:

SEQUENCE CHARACTERISTICS:

LENGTH:48

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:39

GATCAGAGA CATGTCTACC AGACCAGTAC CCGAGCCGG ACGGAATG

48

INFORMATION FOR SEQ ID NO:40:

SEQUENCE CHARACTERISTICS:

LENGTH:35

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:40

GTGATCGTAA CATCCCGCGT TCTCCGACTT CTCGG

35

INFORMATION FOR SEQ ID NO:41:

SEQUENCE CHARACTERISTICS:

LENGTH:36

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:41

CTTCACCGG AGAAGTCGGA GAACGCGGA TGTAC

36

INFORMATION FOR SEQ ID NO:42:

SEQUENCE CHARACTERISTICS:

LENGTH:40

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:42

GGTGAAGGCT TCGTAACTT CGATTACGGT TGGTCGGTG

40

INFORMATION FOR SEQ ID NO:43:

SEQUENCE CHARACTERISTICS:

LENGTH:40

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:43

GTCTGAGCAC CGAACCAACC GTAATCGAAG TTAACGAAGC

40

INFORMATION FOR SEQ ID NO:44

SEQUENCE CHARACTERISTICS:

LENGTH:44

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:44

CTCAGACTGA AGCTGATGCT GATAAGACTG TATGGACCCA TGGA

44

INFORMATION FOR SEQ ID NO:45

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:45

AGCTTCCATG GGTCCATACA GTCTTATCAG CATCAGCTTC A

41

INFORMATION FOR SEQ ID NO:46

SEQUENCE CHARACTERISTICS:

LENGTH:39

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:46

AATTCCCATG GTAACCATTA CCATGCTCCG AACGGTTCT

39

INFORMATION FOR SEQ ID NO:47

SEQUENCE CHARACTERISTICS:

LENGTH:42

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:47

CACCCAGAGA ACCGTTTCGGA GCATGGTAAT GGTTACCATG GG

42

INFORMATION FOR SEQ ID NO:48

SEQUENCE CHARACTERISTICS:

LENGTH:41

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:48

CTGGGTGCTA TGCATGTATA CGAATCTAAA TTCCGTAACT G

41

INFORMATION FOR SEQ ID NO:49

SEQUENCE CHARACTERISTICS:

LENGTH:42

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:49

CTTCAGACCA GTTACGGAAT TTAGATTCGT ATACATGCAT AG

42

INFORMATION FOR SEQ ID NO:50

SEQUENCE CHARACTERISTICS:

LENGTH:37

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:50

GTCTGAAGGT TACTCTGACT TCGATCGTGG TGCTTAC

37

INFORMATION FOR SEQ ID NO:51

SEQUENCE CHARACTERISTICS:

LENGTH:37

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:51

GTGATAACGT AAGCACCACG ATCGAAGTCA GAGTAAC

37

INFORMATION FOR SEQ ID NO:52

SEQUENCE CHARACTERISTICS:

LENGTH:38

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:52

GTTATCACCT TCATTCCGAA ATCTTGAAC ACTGCTCC

38

INFORMATION FOR SEQ ID NO:53

SEQUENCE CHARACTERISTICS:

LENGTH:38

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:53

CTTTGTCCGG AGCAGTGTTC CAAGATTCG GAATGAAG

38

INFORMATION FOR SEQ ID NO:54

SEQUENCE CHARACTERISTICS:

LENGTH:38

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:54

GGACAAAGTT AACAGGGTT GCCCGTAATG AAAGCTTA

38

INFORMATION FOR SEQ ID NO:55

SEQUENCE CHARACTERISTICS:

LENGTH:34

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:55

AGCTTAAGCT TTCATTACGG CCAACCCTGT TTAA

34

INFORMATION FOR SEQ ID NO:56

SEQUENCE CHARACTERISTICS:

LENGTH:20

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:56

TTTTCCTCAGT CACGACGTTG

20

INFORMATION FOR SEQ ID NO:57

SEQUENCE CHARACTERISTICS:

LENGTH:21

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:57

CAGGAAACAG CTATGACCAT G

21

INFORMATION FOR SEQ ID NO:58

SEQUENCE CHARACTERISTICS:

LENGTH:36

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:58

TAAGGAGGTT TAAAATGTCT GACGATCGTG TTACTC

INFORMATION FOR SEQ ID NO:59

SEQUENCE CHARACTERISTICS:

LENGTH:21

TYPE: nucleic acid

STRANDEDNESS:single

TOPOLOGY: linear

MOLECULE TYPE:other nucleic acid synthetic DNA

SEQUENCE DESCRIPTION: SEQ ID NO:59

TACGCCAAGG TTGTTAACCC A

21